

**(19) World Intellectual Property Organization  
International Bureau**



**(43) International Publication Date**  
**25 May 2001 (25.05.2001)**

## PCT

**(10) International Publication Number**  
**WO 01/37571 A1**

**(51) International Patent Classification<sup>7</sup>: H04N 7/24, H04L 12/64**

**(72) Inventors:** RADHA, Hayder; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). LOGUINOV, Dmitri; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

**(21) International Application Number:** PCT/EP00/10902

**(74) Agent: HOEKSTRA, Jelle; Internationaal Octrooibureau B.V., Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).**

**(22) International Filing Date:**  
2 November 2000 (02.11.2000)

**(81) Designated States (national):** AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

**(25) Filing Language:** English

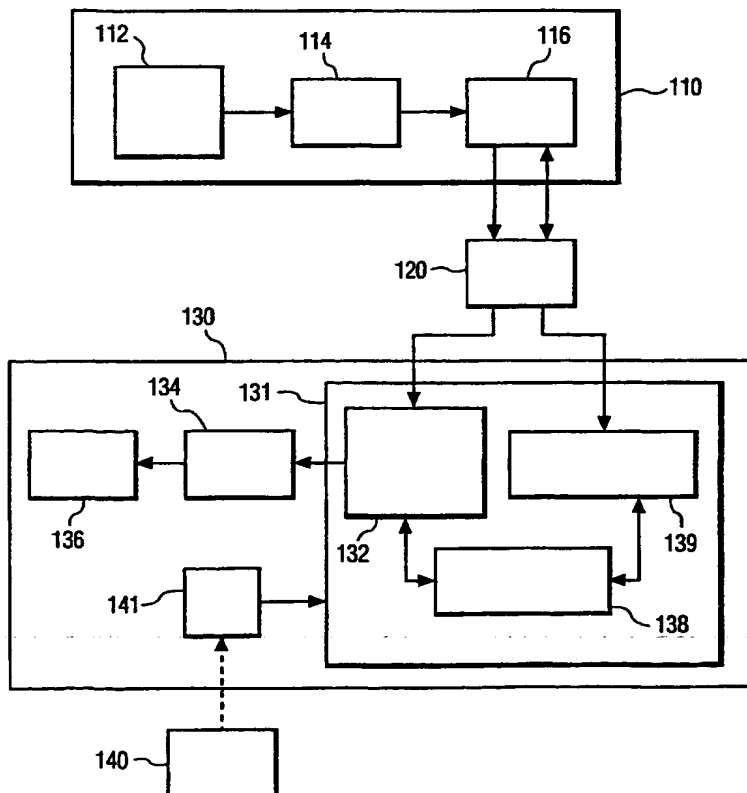
**(26) Publication Language:** English

**(30) Priority Data:**  
09/440,622 15 November 1999 (15.11.1999) US

**(84) Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,

*[Continued on next page]*

**(54) Title: SYSTEM AND METHOD FOR CONTROLLING THE DELAY BUDGET OF A DECODER BUFFER IN A STREAMING DATA RECEIVER**



**(57) Abstract:** There is disclosed a delay budget controller for use with a decoder buffer that receives streaming data packets over a data network from a streaming transmitter and stores the data packets in a plurality of access units for subsequent retrieval by a streaming data decoder. The delay budget controller comprises 1) a first controller for monitoring at least one network parameter associated with the data network; and 2) a second controller for monitoring in the decoder buffer a delay budget region comprising a sequence of access units that are about to be accessed sequentially by the data decoder. The delay budget region comprises a retransmission region and a late region separated by a temporal boundary, wherein the second controller detects missing data packets in the retransmission region and the late region and, in response to detection of a missing data packet in the retransmission region, transmits a retransmission request for the missing data packet to the streaming transmitter. The second controller adjusts the temporal boundary to thereby advance or retard the transmission of the retransmission request.



IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

**Published:**

- *With international search report.*
- *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

System and method for controlling the delay budget of a decoder buffer in a streaming data receiver.

## CROSS-REFERENCE TO RELATED APPLICATIONS

The present invention is related to that disclosed in United States Patent Application Serial No. 09/365,463, entitled "DECODER BUFFER FOR STREAMING VIDEO RECEIVER AND METHOD OF OPERATION," filed on August 2, 1999. The foregoing application is commonly assigned to the assignee of the present invention. The disclosures of the related patent application is hereby incorporated by reference for all purposes as if fully set forth herein.

## TECHNICAL FIELD OF THE INVENTION

The present invention is directed, in general, to data streaming systems and, more specifically, to a decoder buffer for use in a streaming data receiver, such as a streaming video receiver.

## BACKGROUND OF THE INVENTION

Real-time streaming of data, such as multimedia content, over Internet protocol (IP) networks has become an increasingly common application in recent years. A wide range of interactive and non-interactive multimedia Internet applications, such as news on-demand, live TV viewing, video conferencing, and many others, rely on end-to-end streaming solutions. Unlike a "downloaded" audio or video file, which may be retrieved first in "non-real" time and viewed or played back later, streaming audio and streaming video applications require an audio source or video source to encode and to transmit an audio signal or video signal over a network to an audio receiver or a video receiver, which must decode and display (or play back) the transmitted signal in real time. The receiver relies on a decoder buffer to receive encoded video data packets and/or encoded audio data packets from the network and to transfer the packets to a video decoder and/or an audio decoder.

Two problems arise when a streaming data signal is transmitted across a non-guaranteed Quality-of-Service (QoS) network, such as the Internet. First, end-to-end variations in the network (e.g., delay jitter) between the streaming data transmitter and the

streaming data receiver mean that the end-to-end delay is not constant. Second, there is usually a significant packet loss rate across non-QoS networks, often requiring re-transmission. The lost data packet must be recovered prior to the time the corresponding frame must be decoded. If not, an underflow event occurs. Furthermore, if prediction-based compression is used, an underflow due to lost data packets may not only impact the current frame being processed, but may affect many subsequent frames.

It is well-known that re-transmission of lost packets is a viable means of recovery for continuous media communication over packet networks. Many applications use a negative automatic repeat request (NACK) in conjunction with re-transmission of the lost packet. These approaches take into consideration both the round-trip delay and the delay jitter between the sender and the receiver(s).

For example, an end-to-end model with re-transmission for packet voice transmission has been developed. This model takes advantage of the fact that voice data consists of periods of silence separated by brief talk-spurt segments. The model also assumes that each talk-spurt consists of a fixed number of fixed-size packets. However, this model is not general enough to capture the characteristics of compressed video or audio (which can have variable number of bytes or packets per video or audio frame). Additionally, an adaptive playback algorithm has been developed that changes the playback time of a video frame in response to network conditions. This results in a time-varying playback rate (i.e., introduces "playback jitter") in response to network jitter and packet losses.

The above mentioned solutions can be applicable for voice data or for certain video applications which tolerate "playback jitter." However, these solutions may not be acceptable for many types of video-on-demand services (e.g., entertainment applications). In addition, while maintaining continuous decoding and displaying of the real-time audio/visual data, it is crucial for the selected packet loss recovery mechanism to modify its operation according to changing conditions during the Internet session in which the data is transmitted.

Any packet retransmission scheme must strike a balance in determining when to request retransmission of a late data packet. If a streaming data receiver waits too long before requesting retransmission of a late (and possibly lost) data packet, the requested data packet may not be received when needed, due to the round trip delay associated with the retransmission request and retransmission of the late data packet. However, if the streaming data receiver waits only a very brief period before requesting retransmission of a late (but not lost) data packet, an excessive amount of the limited bandwidth available between the streaming data transmitter and the streaming data receiver will be consumed by the increased

number of unnecessary retransmission requests and the increased number of duplicate packet transmissions.

There is therefore a need in the art for improved streaming data receivers that compensate for variations inherent in a non-QoS network. In particular, there is a need for an improved receiver decoder buffer that takes into consideration both transport delay parameters (e.g., end-to-end delay and delay jitter) and video (or audio) encoder buffer constraints. More particularly, there is a need for an improved decoder buffer that implements a packet loss recovery mechanism that modifies its operation according to changing conditions of the data network over which the streaming data is transmitted and minimizes the number of duplicate data packet transmissions.

## SUMMARY OF THE INVENTION

The present invention is embodied in an Integrated Transport Decoder (ITD) buffer model. One key advantage of the ITD model is that it eliminates the separation of a network-transport buffer, which is typically used for removing delay jitter and recovering lost data, from the video/audio decoder buffer. This can significantly reduce the end-to-end delay, and optimize the usage of receiver resources (such as memory).

The present invention provides a re-transmission framework that uses a time-delay budget constraint for streaming video receiver during a real-time Internet session. In other words, at the beginning of the session, the streaming data receiver introduces a certain start-up delay to the incoming bitstream. This start-up delay defines the time-delay budget that the streaming data receiver can rely on for packet loss recovery for the remainder of the session. The re-transmission framework manages this time-delay budget in an adaptive manner in response to changing network conditions. The present invention maximizes the time for uninterrupted decoding and presentation of the multimedia content while minimizing time for duplicate-packet transfer events. These duplicate-packet transfer events occur when the streaming data receiver requests the re-transmission of packets prematurely, reducing the effective available bandwidth between the streaming data transmitter and the streaming data receiver.

It is a primary object of the present invention to provide a delay budget controller for use with a decoder buffer capable of receiving streaming data packets over a data network from a streaming transmitter and storing the data packets in a plurality of access units for subsequent retrieval by a streaming data decoder. In an advantageous embodiment, the delay budget controller comprises 1) a first controller capable of monitoring at least one

network parameter associated with the data network; and 2) a second controller capable of monitoring in the decoder buffer a delay budget region comprising a sequence of access units that are about to be accessed sequentially by the data decoder, the delay budget region comprising a retransmission region and a late region separated by a temporal boundary, wherein the second controller detects missing data packets in the retransmission region and the late region and, in response to detection of a missing data packet in the retransmission region, transmits a retransmission request for the missing data packet to the streaming transmitter, and wherein the second controller is capable of adjusting the temporal boundary to thereby advance or retard the transmission of the retransmission request.

10 In one embodiment of the present invention, the second controller adjusts the temporal boundary in response to a measured value of the at least one network parameter.

In another embodiment of the present invention, the at least one network parameter comprises a round trip delay period associated with the retransmission request.

15 In still another embodiment of the present invention, the at least one network parameter comprises a delay jitter associated with a variation in the round trip delay period.

In yet another embodiment of the present invention, the at least one network parameter comprises an available bandwidth value associated with a communication channel between the streaming data transmitter and the decoder buffer.

20 In a further embodiment of the present invention, the first and second controllers are capable of determining a probability that a packet that is identified as lost by the first and second controllers is actually lost.

In a still further embodiment of the present invention, the second controller adjusts the temporal boundary in response to a value of the probability.

25 In a yet further embodiment of the present invention, the first controller is capable of adjusting a second temporal boundary associated with the delay budget region to thereby increase or decrease a duration of the delay budget region.

Those skilled in the art will readily understand that while the embodiment of the present invention described in the DETAILED DESCRIPTION OF THE INVENTION that follows is principally oriented towards streaming video (or audio), this is by way of illustration only. More broadly speaking, the improved integrated transport decoder buffer described below may be readily adapted for use in connection with any type of streaming data, including video data and audio data, that must be supplied to a decoder at a required rate.

The foregoing has outlined rather broadly the features and technical advantages of the present invention so that those skilled in the art may better understand the detailed description of the invention that follows. Additional features and advantages of the invention will be described hereinafter that form the subject of the claims of the invention.

5 Those skilled in the art should appreciate that they may readily use the conception and the specific embodiment disclosed as a basis for modifying or designing other structures for carrying out the same purposes of the present invention. Those skilled in the art should also realize that such equivalent constructions do not depart from the spirit and scope of the invention in its broadest form.

10 Before undertaking the DETAILED DESCRIPTION, it may be advantageous to set forth definitions of certain words and phrases used throughout this patent document: the terms "include" and "comprise," as well as derivatives thereof, mean inclusion without limitation; the term "or," is inclusive, meaning and/or; the phrases "associated with" and "associated therewith," as well as derivatives thereof, may mean to include, be included  
15 within, interconnect with, contain, be contained within, connect to or with, couple to or with, be communicable with, cooperate with, interleave, juxtapose, be proximate to, be bound to or with, have, have a property of, or the like; and the term "controller" means any device, system or part thereof that controls at least one operation, such a device may be implemented in hardware, firmware or software, or some combination of at least two of the same. It should  
20 be noted that the functionality associated with any particular controller may be centralized or distributed, whether locally or remotely. Definitions for certain words and phrases are provided throughout this patent document, those of ordinary skill in the art should understand that in many, if not most instances, such definitions apply to prior, as well as future uses of such defined words and phrases.

25

## BRIEF DESCRIPTION OF THE DRAWINGS

For a more complete understanding of the present invention, and the advantages thereof, reference is now made to the following descriptions taken in conjunction with the accompanying drawings, wherein like numbers designate like objects, and in which:

30 FIGURE 1 illustrates an end-to-end transmission of streaming video from a streaming video transmitter through a data network to an exemplary streaming video receiver according to one embodiment of the present invention;

FIGURE 2 illustrates an ideal encoder-decoder model of a video coding system;

FIGURE 3 illustrates end-to-end transmission of streaming video from a compressed video source through a channel to an exemplary integrated transport decoder buffer and video decoder, according to one embodiment of the present invention;

FIGURE 4 illustrates a sequence diagram showing the flow of data packets through different and distinct regions of exemplary ideal integrated transport decoder buffer;

FIGURE 5 illustrates an exemplary delay budget controller in accordance with one embodiment of the present invention; and

FIGURE 6 is a flow chart illustrating the operation of an exemplary decoder buffer in accordance with one embodiment of the present invention.

## DETAILED DESCRIPTION OR THE INVENTION

FIGURES 1 through 6, discussed below, and the various embodiments used to describe the principles of the present invention in this patent document are by way of illustration only and should not be construed in any way to limit the scope of the invention.

Those skilled in the art will understand that the principles of the present invention may be implemented in any suitably arranged streaming video receiver.

Additionally, those skilled in the art will readily understand that while the embodiment of the present invention described below is principally oriented towards streaming video, this is by way of illustration only. In fact, the improved integrated transport decoder buffer described below may be readily adapted for use in connection with streaming audio data or other streaming data that must be supplied to a decoder at a required rate.

FIGURE 1 illustrates an end-to-end transmission of streaming video from streaming video transmitter 110 through data network 120 to at least one streaming video receiver 130, according to one embodiment of the present invention. Other streaming video receivers similar to streaming video receiver 130 may also be coupled to streaming video transmitter 110 via data network 120. In order to avoid redundant description, however, these other streaming video receivers are not shown or explained.

Depending on the application, streaming video transmitter 110 may be any one of a wide variety of sources of video frames, including a data network server, a television station, a cable network, a desktop personal computer (PC), or the like. Streaming video transmitter 110 comprises video frame source 112, video encoder 114 and encoder buffer 116. Video frame source 112 may be any device capable of generating a sequence of uncompressed video frames, including a television antenna and receiver unit, a video cassette



player, a video camera, a disk storage device capable of storing a “raw” video clip, and the like.

The uncompressed video frames enter video encoder 114 at a given picture rate (or “streaming rate”) and are compressed according to any known compression algorithm or device, such as an MPEG-4 encoder. Video encoder 114 then transmits the compressed video frames to encoder buffer 116 for buffering in preparation for transmission across data network 120. Data network 120 may be any suitable IP network and may include portions of both public data networks, such as the Internet, and private data networks, such as an enterprise-owned local area network (LAN) or wide area network (WAN).

Streaming video receiver 130 comprises decoder buffer 131, video decoder 134 and video display 136. Decoder buffer 131 receives and stores streaming compressed video frames from data network 120. Decoder buffer 131 then transmits the compressed video frames to video decoder 134 as required. Video decoder 134 decompresses the video frames at the same rate (ideally) at which the video frames were compressed by video encoder 114.

Decoder buffer 131 further comprises integrated transport decoder (ITD) buffer 132, delay budget controller 138 and re-transmission controller 139. In accordance with the principles of the present invention, ITD buffer 132 integrates both temporal and data-unit occupancy considerations in order to provide video decoder 134 with compressed video frames at a rate that is sufficient to avoid underflow conditions, during which video decoder 134 is starved for compressed video frames.

ITD buffer 132 accomplishes this in cooperation with delay budget controller 138 and re-transmission controller 139. Delay budget controller 138 monitors the level of data-occupancy in ITD buffer 132 and detects missing data packets and potential underflow conditions. In response to notification from delay budget controller 138, re-transmission controller 139 requests re-transmission of data missing from ITD buffer 132 in order to avoid underflow conditions.

In an advantageous embodiment of the present invention, ITD buffer 132, delay budget controller 138, and re-transmission controller 139 are implemented in a personal computer (PC) that receives streaming video and/or audio from, for example, the Internet over a high-speed data line. In such an embodiment, ITD buffer 132 may be implemented in main random access memory (RAM) of the PC or in RAM on a video card, and delay budget controller 138 and re-transmission controller 139 may be implemented in the CPU of the PC. To implement delay budget controller 138 in a PC environment, the

present invention may be embodied as computer executable instructions stored as a program on storage media 140. Storage media 140 may be a CD-ROM, a computer diskette, or a similar device that may be loaded into removable disk port 141 in streaming video receiver 130.

5 Continuous decoding of compressed video frames is a key requirement of a real-time multimedia application, such as streaming video. To meet this requirement, a decoder-encoder buffer model is normally used to ensure that underflow and overflow events do not occur. These constraints limit the size (bit-wise) of video pictures that enter the encoder buffer. The constraints are usually expressed in terms of encoder-buffer bounds,  
10 which when adhered to by the encoder, guarantee continuous decoding and presentation of the compressed video stream at the receiver.

FIGURE 2 shows an ideal encoder-decoder model of a video coding system. Under this ideal model, uncompressed video frames 201-203 enter the compression engine of encoder 214 at a given picture-rate,  $X$  frames/second, as indicated by the Time(1) line. The  
15 compressed frames exit encoder 214 and enter encoder buffer 216 at the same  $X$  frames/second, as indicated by the Time(2) line. Similarly, the compressed frames exit decoder buffer 216 and enter channel 220 at  $X$  frames/second. Channel 220 is a generic representation of any transmission medium, such as the Internet, that transfers compressed video frames from a transmitting source to a receiver. In the ideal case, the delay of  
20 channel 220 ( $\delta_C$ ) is a constant value.

Next, the compressed frames exit channel 220 and enter decoder buffer 232 at the same  $X$  frames/second as at the input and the output of encoder 214, as indicated by the Time(3) line. Decoder buffer 232 transmits the compressed frames to decoder 234, which decompresses the frames and outputs decompressed frames 251-253 at the original  $X$   
25 frames/second at which frames entered encoder 214.

Ideally, the end-to-end buffering delay (i.e., the total delay encountered in both encoder buffer 216 and decoder buffer 232) is constant. However, the same piece of compressed video data (e.g., a particular byte of the video stream) encounters different delays in encoder buffer 216 and decoder buffer 232. In the ideal model, encoding in encoder 214  
30 and decoding in decoder 234 are instantaneous and require zero execution time and data packets are not lost.

The encoder buffer bounds can be expressed using discrete-time summation. In discrete-time domain analysis,  $\Delta$  is the end-to-end delay (i.e., including both encoder buffer 216 and decoder buffer 232 and channel  $\delta_C$ ) in units of time. For a given video coding

system,  $\delta$  is a constant number applicable to all frames entering the encoder-decoder buffer model.

To simplify the discrete-time analysis, it is assumed that the end-to-end buffering delay ( $\Delta T = \Delta - \delta_c$ ) is an integer-multiple of the frame duration ( $T$ ). Therefore,  $N\Delta =$   
 5  $N(\Delta - \delta_c)/T$  represents the delay of the encoder and decoder buffers in terms of the number of video frames ( $N$ ). For the purposes of clarity and brevity in describing the principles of the present invention, the remainder of this disclosure will use time units specified in frame-duration intervals. For example, using the encoder time reference shown in FIGURE 2, the  $n^{\text{th}}$  frame enters encoder buffer 216 at time index “ $n$ ”. The decoder time-reference of decoder  
 10 buffer 232 is shifted by the channel delay ( $\delta_c$ ), with respect to encoder buffer 216.

The data rate ( $r$ ) at the output of encoder (e) 214 during frame-interval “ $i$ ” may be represented as  $r^e(i)$ . Here, “data rate” is used generically. It could signify bit rate, byte rate, or even packet rate. Similarly, the data rate at the input of decoder buffer 232 may be represented as  $r^d(i)$ . Based on the ideal model,  $r^e(iT) = r^d(iT + \delta_c)$ . In addition, based on the  
 15 convention established above,  $r^e(i) = r^d(i)$ . Thus, the bounds of encoder buffer 216 can be expressed as:

$$\max \left[ \left( \sum_{j=n+1}^{n+\Delta N} r^e(j) - B_{\max}^d \right), 0 \right] \leq B^e(n) \leq \min \left[ \left( \sum_{j=n+1}^{n+\Delta N} r^e(j) \right), B_{\max}^e \right] \quad [\text{EQUATION 1}]$$

20 where  $B_{\max}^d$  and  $B_{\max}^e$  are the maximum decoder and encoder buffer sizes respectively.

In the ideal case, it is assumed that encoder 214 begins to transmit data immediately after the first frame enters encoder 214. Therefore, the start-up delay  $dd_f$  (i.e., the delay time the first piece of data from the first picture spends in decoder buffer 232 prior to decoding) equals the end-to-end, encoder-decoder buffer delay:  $dd_f = \Delta T = T \cdot \Delta N$ .

25 In one embodiment of the present invention, ITD buffer 132 minimizes underflow events by taking into consideration the above-described problems of the ideal buffer model and the ideal encoder-decoder buffer constraints. ITD buffer 132 is based on lost packet recovery using re-transmission.

FIGURE 3 is a simplified block diagram of exemplary end-to-end  
 30 transmission of streaming video, without support for re-transmission. For the purposes of simplicity and clarity, streaming video transmitter 110 has been replaced by compressed video source 305 and data network 120 has been replaced by channel 320. Compressed video

source 305 transmits data packets at rate  $r^e(n)$  and channel 320 transmits data packets at rate  $r^{td}(n)$ . Since video re-transmission is not discussed in FIGURE 3, delay budget controller 138 and re-transmission controller 139 are omitted from the diagram. Streaming video receiver 130 has been simplified and is represented by ITD buffer 132 and video decoder 134.

As noted above, ITD buffer 132 integrates temporal and data-unit occupancy models. ITD buffer 132 is divided into temporal segments of 'T' seconds each. By way of example, the parameter T may be the frame period in a video sequence. The data packets (bits, bytes, or packets) associated with a given duration T are buffered in the corresponding temporal segment. All of the data packets associated with a temporal unit are referred to as an "access" unit. By way of example, data packets 351, 352, and 353 comprise access unit  $A_{n+1}$ , data packet 354 comprises access unit  $A_{n+2}$ , and data packets 355 and 356 comprise access unit  $A_{n+3}$ .

During time interval n, the  $n^{th}$  access unit,  $A_n$ , is being decoded by decoder 134 and access unit  $A_{n+1}$  is stored at the temporal segment nearest to the output of ITD buffer 132. An access unit may be an audio frame, a video frame, or even a portion of a video frame, such as Group of Blocks (GOB). Therefore, the duration required to decode or to display an access unit is the same as the duration of the temporal segment T. During the time-interval n, the rate at which data enters ITD buffer 132 is  $r^{td}(n)$ . The numbers of data packets in each access unit are not required to be the same. Compression algorithms used in video encoder 114 may compress the data packets in successive access units by different amounts, even though each access unit represents temporal units of the same duration.

For example, the three data packets 351-353 in access unit  $A_{n+1}$  may comprise a complete video frame, Frame 1. The single data packet 354 in  $A_{n+2}$  may represent only those portions of Frame 2 that are different than Frame 1. Nonetheless, data packet 354 is sufficient to create Frame 2 if the Frame 1 data is already known. Since Frame 1 and Frame 2 have the same duration, the temporal segment, T, is the same for  $A_{n+1}$  and  $A_{n+2}$ .

Each temporal segment holds a maximum number of packets,  $K_{max}$ , with each packet having a maximum size,  $b_{max}$ , (in bits or bytes). Therefore, the maximum size of an access unit,  $S_{max}$ , may be represented by  $S_{max} \leq K_{max}(b_{max})$ . Video encoder 114 is assumed to begin each access-unit with a new packet that is present only in that access unit.

The amount of data in ITD buffer 132 at time index n,  $B^{td}(n)$ , may be described by terms of  $B^a(n)$  and  $B^b(n)$ .  $B^a(n)$  represents the number of consecutive-and-complete access units in ITD buffer 132 at the beginning of interval n, and  $B^b(n)$  represents the total consecutive amount of data in ITD buffer 132 at the end of interval n. For  $B^a(n)$ ,

temporal segments containing partial data are not counted, and all segments following a partial segment are also not counted even if they contain a complete, access-unit worth of data. Hence,  $TB^a(n)$  represents the amount of video data in temporal units (e.g., seconds) that ITD buffer 132 holds at time index  $n$  (without running into an underflow if no more data arrives).

If  $S_n$  denotes the size of access unit  $n$ , the relationship between  $B^a$  and  $B^b$  can be expressed as Equation 2 below:

$$B^b(n) = \sum_{j=N+1}^{n+B^a(n)} S_j + U_{B^a(n)+1} \quad \text{[EQUATION 2]}$$

where  $S_j$  is the maximum size of the access unit for temporal segment  $j$  and  $U_{B^a(n)+1}$  is the partial (incomplete) data of access unit  $A_{n+B^a(n)+1}$  stored in temporal segment  $B^a(n)+1$  at the beginning of time index  $n$ .

In order for re-transmission to occur, ITD buffer 132 must be able: a) to output one temporal segment (T) worth of data at the beginning of every temporal time-interval  $n$ ; b) to detect lost packet(s) and to transmit corresponding negative acknowledge (NACK) messages to transmitter 110 or compressed video source 305; c) to continuously store newly arrived primary (i.e., not re-transmitted) packets; and d) to store re-transmitted packets. Ideal ITD buffer 132 maintains the data rate of the video stream, without delays caused by re-transmission of any lost data. In other words, if  $r^e(n)$  is the transmission data rate used by idealized video encoder 114 under lossless circumstances, ideal ITD buffer 132 will maintain this data rate without degradation caused by the re-transmission process. Depending upon the number of re-transmission requests, encoder buffer 116 may adjust its output data rate  $r^e(n)$ , with a corresponding adjustment by ITD buffer 132.

In one embodiment, decoder buffer 131 adds buffering for the incoming video stream in order to compensate for the time required for detection and recovery of lost data and for the delay associated with a "real" world implementation. By delaying all incoming video streams by this compensation time, decoder buffer 131 outputs video stream data at a continuous rate as required for decoding. The minimum duration of time needed for detecting a predetermined number of lost packets is represented by  $T_L$ . In general,  $T_L$  is a function of the delay jitter caused by data arriving later than expected by ITD buffer 132.

The minimum amount of time needed for streaming video receiver 130 to recover a packet after being declared lost is represented by  $T_R$ . Time  $T_R$  includes the time required for streaming video receiver 130 to send a NACK to streaming video transmitter 110

and the time needed for the re-transmitted data to reach streaming video receiver 130 (assuming that the NACK and re-transmitted data are not lost).

Exemplary decoder buffer 131 transfers a re-transmitted packet with a minimum delay ( $T_L + T_R$ ) for the lost packet interval. If the minimum delay experienced by any video data for an ideal decoder buffer 131 is represented by  $dd_{min}$ , the amount of delay that may be added to the minimum ideal delay in order to account for the total delay for re-transmission is:

$$\Delta_R \geq u(T_L + T_R - dd_{min}) \quad [\text{EQUATION 3}]$$

where  $u(x)=x$  for  $x>0$ , and  $u(x)=0$  for  $x \leq 0$ .

Decoder buffer 131 adds delay  $\Delta_R$  buffering for all output data to video decoder 134 in order to provide time for decoding and transferring of the data, resulting in continuous video streams. Therefore, the total encoder buffer 116 to decoder buffer 132 output delay ( $\Delta_{TOT}$ ) may be represented by:

$$\Delta_{TOT} = \Delta_{ideal} + \Delta_R \geq \Delta_{ideal} + u(T_L + T_R - dd_{min}) \quad [\text{EQUATION 4}].$$

ITD buffer 132 provides buffering (storage) for a minimum number of temporal segments ( $B_{min}^a$ ) as compensation for re-transmission time requirements and as prevention for an underflow event. The ITD buffer 132 sizing may be based, for example, on minimum and maximum boundaries for storing temporal segments. The process for determining these boundaries was described in United States Patent Application Serial No. 09/365,463, entitled "DECODER BUFFER FOR STREAMING VIDEO RECEIVER AND METHOD OF OPERATION" and previously incorporated by reference into this disclosure.

FIGURE 4 is a sequence diagram showing the flow of data packets through different regions of exemplary ITD buffer 132 under the assumption that  $dd_{min} = 0$  (the lower boundary level) and  $d_{max} = \Delta_{ideal}$ . ITD buffer 132 data enters from the right side of the diagram and exits to the video decoder 134 at the left side. The most recently received data is in a buffer area which is labeled "too-early for re-transmission request region" (too-early). Depending on the location in the too-early region of the buffer, ITD buffer 132 introduces buffer delays labeled  $N_E$ ,  $\Delta N$ , or  $N_L$ . The area of this too-early buffer region, which comprises the ideal delay  $\Delta N$ , is labeled as the ideal-buffer region. ITD buffer 132 manages the ideal-buffer region as an ideal video buffer. In other words, data packets flow through this region and are only delayed by the inherent characteristics of the buffer element(s). Ideal ITD buffer 132 provides the remaining too-early buffer areas to compensate for delays associated with

the transfer of video streams from streaming video transmitter 110 to decoder 131 ( $N_E$ ), as well as delays caused by delayed or lost video packets ( $N_L$ ).

ITD buffer 132 provides a delay,  $N_R$ , in the re-transmission region in order to compensate for expected time requirements for the initiation and reception of re-transmission requests. Exemplary decoder buffer 131 initiates re-transmission requests during the time periods associated with the re-transmission region.

It is important to note that the ideal-buffer and re-transmission regions may overlap, depending on the values of the different delay parameters ( $dd_{min}$ ,  $T_R$ ,  $T_L$ ). However, for the exemplary ideal ITD buffer 132 with  $dd_{min}=0$ , the re-transmission and ideal-buffer regions do not overlap.

For ITD buffer 132,  $N_E$  represents the initial decoding delay ( $dd_f$ ) which corresponds to the amount of delay encountered by the very first piece of data that enters the buffer prior to the decoding of the first picture (or access unit). This  $dd_f$  is based on, among other things, the streaming video transmitter 110 and data network 120 data transmission rates during elapsed time  $dd_f$ . In the ideal case, ITD buffer 132 uses this same data rate for entering received data into its buffer (storage) regions. Ideal decoder buffer 131 recognizes the amount of data in its ITD buffer 132 regions just prior to the time that the first access unit is decoded as  $B_0^d$  data. The  $B_0^d$  data, also referred to as "start-up-delay" data, is given by the following relationship:

$$B_0^d = \sum_{j=1}^{\Delta N} r^e(j) \quad \text{[EQUATION 5].}$$

When  $dd_{min}=0$ , ideal decoder buffer 131 re-transmission processing is comprised of the following procedures:

1. The ideal-buffer region is filled until all data associated with the start-up delay are in the buffer. Since lost events may also occur during this time interval, these data may be treated in a special way, such as by using reliable transmission (e.g., using TCP) for them. The ideal condition for lossless data is satisfied when:

$$\sum_{k=N_R+N_L+1}^{N_R+N_L+\Delta N} B_k = B_0^d \quad \text{[EQUATION 6]}$$

where  $B_k$  is the amount of data stored in ideal ITD buffer 132 temporal segment  $k$  at any instant of time.

2. After Equation 15 is satisfied, ITD buffer 132 advances the content of all temporal storage segments by one segment toward the buffer output. Subsequently, ideal ITD

buffer 132 repeats this process every  $T$  units of time. After  $N_L + N_R$  periods of  $T$  (i.e., after  $T_L + T_R$ ), decoder 134 starts decoding the first access unit. The time-period that starts when decoding of the first access unit begins is labeled  $T_1$ . Hence, the beginning of any time period  $n$  ( $T_n$ ) represents the time when access unit  $A_{n+k}$  is moved to temporal segment  $k$ .

- 5                    3. Ideal ITD buffer 132 considers data missing in temporal segment  $N_R$  of the re-transmission buffer region as lost. This condition occurs when:

$$B_{N_R}(n) < S_{n+N_R} \text{ [EQUATION 7]}$$

where  $B_{N(R)}$  is the amount of data in temporal segment  $N_R$  at time period  $n$  and  $S_j$  is the size of access unit  $j$ . When ideal ITD buffer 132 determines that data is missing, it sends a re-  
10                    transmission request to streaming video transmitter 110.

4. Ideal ITD buffer 132 places arriving re-transmitted data into their corresponding temporal segments of the re-transmission region. Assuming the re-transmitted data are received, ideal ITD buffer 132 transfers the re-transmitted data to the video decoder 134 prior to the decoding times of their corresponding access units.

- 15                    In order to properly balance the need to request retransmission of late packets in a timely manner with the need to avoid duplicative data packet deliveries, the present invention provides a retransmission framework that uses a time-delay budget constraint for streaming data receivers during real-time Internet sessions. Delay budget controller 138 manages this time-delay budget in an adaptive manner in response to changing network  
20                    conditions, as described below in greater detail.

The present invention is primarily receiver-oriented retransmission technique. FIGURE 5 illustrates exemplary delay budget controller 138 in accordance with one embodiment of the present invention. Delay budget controller 138 comprises real-time QoS characterization circuit 505, buffer management circuit 510, and delay budget management  
25                    circuit 515, which coordinate the flow of streaming video between streaming video receiver 130 and streaming video transmitter 110.

In one embodiment of the present, delay budget controller 138 establishes a delay budget that is static and adapts the sizes of different temporal regions within the static delay budget. In an alternate embodiment of the present invention, delay budget controller  
30                    138 may establish a delay budget that is dynamic and may adapt the size of the over all delay budget, as well as adapt the size of the different temporal regions within the dynamic delay budget.



This is illustrated in FIGURE 4. The temporal position of one or both of boundary lines 405 and 410 may be adjusted by delay budget controller 138, as indicated by the double arrow lines 406 and 411. In a static mode, delay budget controller 138 does not adjust the position of boundary line 410, so that the delay budget =  $N_R + N_L$  is constant. For example, if  $N_R$  is initially selected so that buffer region corresponding to  $N_R$  is 2 seconds and  $N_L$  is selected so that the buffer region corresponding to  $N_L$  is 0.5 seconds, then the delay budget is held constant at 2.5 seconds by delay controller 138. Delay controller 138 may still adjust the position of boundary line 405 within the delay budget, but the total value of the delay budget remains fixed at 2.5 seconds. In a dynamic mode, delay budget controller 138 may adjust the position of boundary line 410, so that the delay budget comprises  $N_R$ ,  $N_L$ , and at least part of  $\Delta N$ . Delay budget controller 138 may also adjust the position of boundary line 405 in the dynamic mode.

Real-time QoS characterization circuit 505 continuously calculates and updates the characterization of Internet sessions in real-time using three network parameters: round-trip delay ( $\delta_{rt}(t)$ ), delay jitter ( $\delta_{jit}(t)$ ), and bandwidth ( $\beta(t)$ ). An exemplary network-based random process model  $\aleph$  may then be derived based on a set of observations,  $\Theta(t) = \{\delta_{jit}(t), \delta_{rt}(t), \beta(t)\}$ , using these three QoS parameters. Once calculated, real-time QoS characterization circuit 505 communicates the network QoS model set for a particular time  $t$ ,  $\aleph(t; \Theta)$ , to other real-time processes in streaming video receiver 130:

$$\aleph(t; \Theta) = \{\delta_{jit}(t), \delta_{rt}(t), \beta(t)\} \quad \text{[EQUATION 8]}$$

Although real-time QoS characterization circuit 505 primarily determines QoS characterization parameters, assistance with these calculations may occasionally be needed from streaming video transmitter 110.

Buffer management circuit 510 identifies internal and external delay times and manages the flow of data for ITD buffer 132. At the beginning of each streaming session, buffer management circuit 510 identifies two minimum delay parameters for buffering media streams. The delay parameters are the minimum ideal start-up delay,  $\Delta_{su}$ , and the minimum total start-up delay,  $D_{su}$ . The value  $\Delta_{su}$  represents the minimum buffering delay needed to ensure that a multimedia video clip will not encounter an underflow condition during the streaming session (assuming no network delay jitter). Since  $\Delta_{su}$  is a function of the particular media stream, its value can be communicated to buffer management circuit 510 by different means. For instance,  $\Delta_{su}$  information may be carried in the media stream, as in the case with the MPEG-2 transport stream and PES layer time stamps.

Buffer management circuit 510 may identify or compute the value for  $D_{su}$  in several ways. One criterion is that  $D_{su}$  should always be larger than  $\Delta_{su}$ . In one embodiment, buffer management circuit 510 may obtain the value for  $D_{su}$  from initialization data which may be modified by the user. In another embodiment, buffer management circuit 510 may access a programmable or user entered maximum start-up delay which can be substituted for  $D_{su}$  or used as the basis for calculating  $D_{su}$ .

In another embodiment, buffer management circuit 510 may calculate preliminary estimates for the round-trip delay  $\delta_{rtt}(t)$ , delay jitter  $\delta_{jit}(t)$ , and the bandwidth  $\beta(t)$ , while ITD buffer 132 is storing streaming data during the initial ideal  $\Delta_{su}$  period. Buffer management circuit 510 may then use these QoS parameters for calculating a desirable value for  $D_{su}$  based on some criterion (e.g., pick a minimum  $D_{su}$  value that enables decoder buffer 131 to recover lost packets with high probability of success). An exemplary buffer management circuit 510 computation for  $D_{su}$  based on preliminary QoS characterization estimates is described below. Alternatively, buffer management circuit 510 may use a combination of user/application driven or network-condition driven applications, or other approaches, for computing  $D_{su}$ .

After  $\Delta_{su}$  and  $D_{su}$  have been identified, budget management circuit 510 calculates the minimum delay budget ( $D_{budget}$ ), where  $D_{budget} = D_{su} - \Delta_{su}$ . The  $D_{budget}$  represents the minimum amount of time available for detecting and recovering lost packets for the remainder of the data streaming session.

Once initialized and after receiving required time allocations for detecting and recovering lost packets, buffer management circuit 510 monitors ITD buffer 132 for the absence of expected packets. A minimum monitoring detection time ( $\tau_{det}$ ) determines the time before a NACK is sent to streaming video transmitter 110. Buffer management circuit 510 may transfer from 1 to R NACKS for each temporal segment of "T" seconds, where R represents the number of access units present in T.

Delay-budget management circuit 515 determines and manages the elapsed time needed for detecting and for recovering lost packets through the re-transmission process. In one embodiment, delay-budget management circuit 515 accomplishes these responsibilities by allocating portions of the time calculated for a static  $D_{budget}$  as packet loss detection time ( $\tau_{det}$ ) and as packet re-transmission time ( $\tau_{rtx}$ ) so that  $D_{budget} = \tau_{det} + \tau_{rtx}$ .

Delay-budget management circuit 515 takes into consideration the real-time network condition model  $\aleph$  derived by real-time QoS characterization circuit 505:

$$\tau_{\text{det}}(t; \mathbf{x}) + \tau_{\text{rtx}}(t; \mathbf{x}) = D_{\text{budget}} \text{ [EQUATION 9]}$$

The above equation expresses the constraints that delay-budget management 515 circuit must adhere to when using a static value for  $D_{\text{budget}}$ . A generalization of this static approach to a time-varying (or dynamic) delay-budget framework is described below.

5 There are a number of ways of determining  $\tau_{\text{det}}$  and  $\tau_{\text{rtx}}$ . In one embodiment of the present invention, during a session delay-budget management circuit 515 continuously adapts to changing network conditions by generating the current estimates of positive delay jitter  $\Delta_{\text{jit}}(t)$  and the round-trip delay  $\Delta_{\text{rtt}}(t)$  using the set of observations  $\Theta(t) = \{\delta_{\text{jit}}(t), \delta_{\text{rtt}}(t)\}$  that were collected in real time. Once estimates of  $\Delta_{\text{jit}}(t)$  and  $\Delta_{\text{rtt}}(t)$  are determined, delay-  
10 budget management circuit 515 may scale these values to fit into the delay budget and assigns the scaled values to  $\tau_{\text{det}}(t)$  and  $\tau_{\text{rtx}}(t)$  as follows:

$$\tau_{\text{rtx}}(t) = [w_{\text{rtt}}(t)\Delta_{\text{rtt}}(t)/(w_{\text{rtt}}(t)\Delta_{\text{rtt}}(t)+w_{\text{jit}}(t)\Delta_{\text{jit}}(t))]D_{\text{budget}}$$

and

$$\tau_{\text{det}}(t) = [w_{\text{jit}}(t)\Delta_{\text{jit}}(t)/(w_{\text{rtt}}(t)\Delta_{\text{rtt}}(t)+w_{\text{jit}}(t)\Delta_{\text{jit}}(t))]D_{\text{budget}}$$

15 For a given pair of estimates,  $\Delta_{\text{rtt}}(t)$  and  $\Delta_{\text{jit}}(t)$ , if the weighting factors  $w_{\text{rtt}}(t)$  and  $w_{\text{jit}}(t)$  are equal, these formulas give equal weight to both the packet-loss detection process and the retransmission process. Therefore, the delay budget,  $D_{\text{budget}}$ , is utilized in a “fair” manner between the two processes. This is accomplished by keeping the ratio  $\tau_{\text{rtx}}(t)/\tau_{\text{det}}(t)$  the same as  $\Delta_{\text{rtt}}(t)/\Delta_{\text{jit}}(t)$  for all times  $t$ .

20 Delay-budget management circuit 515 uses a packet-lost time allocation process that minimizes the time needed for declaring a packet lost, while providing enough time so that the probability that the packet is really lost is as high as possible. The latter criterion is important for minimizing the duplicate transfer of packets and increasing the effective available bandwidth over the Internet session under consideration.

25 Once the packet-lost time ( $\tau_{\text{det}}$ ) has been allocated, delay-budget management circuit 515 allocates the remaining time of  $D_{\text{budget}}$  as time for recovering the lost packet. Thus, the time allocated for detecting a lost packet is based on minimizing the time for detecting a lost packet, while maximizing the available time for recovering the lost packet. When successful, the lost packet is recovered prior to the time when it is needed by video decoder  
30 134.

In some cases, streaming video receiver 130 is constrained to operate within a maximum start-up delay due, for example, to user/application requirements. In these cases,

buffer management circuit 510 treats the determination of  $D_{su}$  as trivial and only considers the minimum (ideal) start-up delay  $\Delta_{su}$  in its calculations.

When buffer management circuit 510 is not constrained by a strict maximum start-up delay, it may limit the time for the start-up delay in order to improve, for example, the system response time to the user or to use system resources (e.g., memory) more efficiently. In one embodiment, buffer management circuit 510 estimates start-up delay using a probabilistic approach based on generating preliminary models for the QoS parameters. These preliminary models may be generated by buffer management circuit 510 during the initial (ideal)  $\Delta_{su}$ , with consideration given to prior knowledge or assumptions about the characteristics of the Internet session under consideration. In addition, buffer management circuit 510 bases its estimate for  $D_{su}$  on a desirable success ratio for packet loss recovery and a desirable confidence level in accurately identifying packet-loss occurrences.

For this embodiment,  $P_R$  represents the desirable probability for successfully recovering a lost packet (i.e., recovering the packet prior to the time it is needed by the decoder). Similarly,  $P_L$  represents the probability that a packet has been lost. Based on the ITD buffer model introduced in United States Patent Application Serial No. 09/365,463, buffer management 510 computes  $D_{su}$  as follows:

$$D_{su} = P_{rt}^{-1}(P_R) + P_{jit}^{-1}(P_L) + \Delta_{su} \quad [\text{EQUATION 10}]$$

where  $P_{rt}(\delta)$  and  $P_{jit}(\delta)$  are the probability distribution functions of the round-trip delay and the positive delay-jitter parameters, respectively.  $P_{rt}^{-1}(\cdot)$  and  $P_{jit}^{-1}(\cdot)$  are the corresponding inverse functions for  $P_{rt}(\delta)$  and  $P_{jit}(\delta)$ .

Using these probability functions, buffer management circuit 510 measures positive-jitter events only for packets arriving later than expected. Consequently, buffer management circuit 510 derives the positive-jitter density-function  $p_{jit}(\delta)$  by normalizing the overall jitter probability-density function:

$$p_{jit}(\delta) = p_{jit}(\delta) / [1 - P_{jit}(0)] \quad [\text{EQUATION 11}]$$

For these calculations, buffer management circuit 510 assumes that  $\delta_{rt}$  represents the total time for recovering a lost packet or the time interval from when the packet is declared lost through the re-transmission time. Moreover, buffer management circuit 510 assumes that the positive delay-jitter as received from data network 120 represents the time needed for declaring a packet lost ( $\delta_{jit}(t)$ ).

As mentioned previously, delay-budget management circuit 515 ensures a high confidence level for determining the amount of time allocated for identification of lost packets in order to minimize the transfer of duplicate packets. This high confidence level is especially needed when the available bandwidth is virtually 100% occupied by the multimedia stream being transmitted.

However, in cases when the available bandwidth  $\beta(t)$  is higher than the bitstream bit rate  $r(t)$ , delay-budget management circuit 515 may utilize some of the excess bandwidth  $[\beta(t)-r(t)]$  for increasing the probability of recovering lost packets. In these cases, delay-budget management circuit 515 may decrease the value for  $\delta_{jit}(t)$  which may cause an increase of duplicate packet transmissions. When excess bandwidth is present, these duplicated packets will not compete with the primary bitstream. The extra time available for re-transmission of lost-packets may improve the overall packet-loss-recovery mechanism without causing a significant increase in the packet loss ratio.

In one embodiment, delay-budget management circuit 515 enhances the ability to calculate a minimum value for  $\delta_{jit}$  across time  $t$  by considering the change in excess bandwidth across time and the desirable probability for successful recovery of a lost packet.

Delay-budget management circuit 515 determines that a packet is lost when the following equation is satisfied:

$$P_L = \int_0^{P_{jit}^{-1}(P_L)} p_{jit}(\delta; t) d\delta \quad [\text{EQUATION 12}]$$

where  $P_L$ , in this case, represents the probability that a packet has been lost when it is declared lost.

In other words, when the probability of a packet-loss event for a particular packet reaches  $P_L$ , delay-budget management circuit 515 declares the packet lost by the system. In a static implementation for  $P_L$ , delay-budget management circuit 515 simply sets  $P_L$  to equal the desired static confidence level.

Delay-budget management circuit 515 modifies the value of  $P_L$  in response to the network condition and to the characteristic of the stream bit rate. In one embodiment, delay-budget management circuit 515 quantifies the change of  $P_L$  across time ( $P_L(t)$ ) as follows:

$$P_L(t) = \int_0^{P_{jit}^{-1}(P_L(t))} p_{jit}(\delta; t) d\delta \propto \frac{r(t)}{\beta(t)} \quad [\text{EQUATION 13}]$$

Using this relationship, delay-budget management circuit 515 chooses a smaller  $P_L$  as the trigger for a packet-loss declaration event when the available bandwidth is high.

As explained above, in the absence of any positive delay-jitter in the network, the delay-budget management circuit 515 may rely on the minimum delay  $D_{\text{budget}}$  for packet loss recovery. This is due to the fact that the ideal decoding delay of the media stream can be as small as zero during any given session. For example, the Integrated Transport Decoder (ITD) buffer model described in United States Patent Application Serial No. 09/365,463 incorporates an ideal decoder section that can be drained even in the absence of network jitter or packet loss events. However, this ideal decoder section can include a delay  $\Delta_{\text{dec}}(t)$  that varies in time and can be as large as the end-to-end encoder-decoder ideal buffering delay  $\Delta$ .

For the static embodiment of delay-budget management circuit 515, delay-budget management circuit 515 only focuses on and manages the re-transmission section of ITD buffer 132. This embodiment provides the means for simplifying ITD buffer 132 and the time-delay management process.

In another embodiment which incorporates a relatively small increase in complexity, streaming video receiver 130 uses a dynamic time-varying delay budget for its packet loss recovery:

$$D(t) = D_{\text{budget}} + \Delta_{\text{dec}}(t) \text{ [EQUATION 14]}$$

where  $D_{\text{budget}}$  is the same parameter as briefly described for the static embodiment of delay-budget management 515. One dynamic embodiment of delay-budget management circuit 515 continuously monitors the ideal-buffer section in order to estimate  $\Delta_{\text{dec}}(t)$ . Furthermore, the delay-budget constraint expressed in EQUATION 8 is modified as follows:

$$\tau_{\text{det}}(t; \mathbf{x}) + \tau_{\text{rx}}(t; \mathbf{x}) = D(t) \text{ [EQUATION 15]}$$

where:

$$D_{\text{budget}} \leq D(t) \leq D_{\text{budget}} + \Delta \text{ [EQUATION 16]}$$

This dynamic framework can be further generalized to include multiple attempts for packet loss recovery. In this case, depending on the amount of delay available, buffer management circuit 510 can request  $n$  packet re-transmission times:

$$n(t) [\tau_{\text{det}}(t; \mathbf{x}) + \tau_{\text{rx}}(t; \mathbf{x})] = D(t) \text{ [EQUATION 10]}$$

Therefore, as the occupancy of ITD buffer 132 changes, the buffer management circuit 510 can change  $n$ . This capability approaches the benefits achieved from the instantaneous delay

$\Delta_{\text{dec}}(t)$  experienced in the ideal-buffer section of the ITD buffer model introduced in United States Patent Application Serial No. 09/365,463.

FIGURE 6 is a flow chart illustrating the operation of exemplary decoder buffer 131 in accordance with one embodiment of the present invention. Initially, delay budget management circuit 515 identifies a minimum total start-up delay ( $D_{\text{SU}}$ ) and a minimum ideal start-up delay ( $\Delta_{\text{SU}}$ ), and calculates a delay budget ( $D_{\text{BUDGET}}$ ) (process step 605). Real-time QoS characterization circuit 505 monitors conditions on data network 120 and collects measured data for round-trip delay,  $\delta_{\text{rt}}(t)$ , delay jitter,  $\delta_{\text{jit}}(t)$ , and bandwidth,  $\beta(t)$  (process step 610).

Delay budget controller 138 monitors ITD buffer 132 to detect the absence of expected packets. If a packet is not received after a minimum monitoring time, it is determined to be lost and a NACK message is sent to streaming video transmitter 110 (process step 615). Using the statistics gathered by real-time QoS characterization circuit 505, delay budget controller 138 calculates the probability distribution functions for round trip delay and delay jitter. Delay budget controller 138 also calculates the probability that a video packet has actually been lost when it was declared lost (process step 620).

If the probability that a particular data packet has been lost exceeds an established probability threshold, delay budget controller 138 transmits a retransmission request to streaming video transmitter 110 (process step 625). Optionally, delay budget controller 138 may adjust the value of the established probability threshold in order to adapt to changing delay conditions on network 120 (process step 630).

Although the present invention has been described in detail, those skilled in the art should understand that they can make various changes, substitutions and alterations herein without departing from the spirit and scope of the invention in its broadest form.

## CLAIMS:

1. For use with a decoder buffer (131) capable of receiving streaming data packets over a data network (120) from a streaming transmitter (110) and storing said data packets in a plurality of access units for subsequent retrieval by a streaming data decoder (134), a delay budget controller (138) comprising:

5 a first controller (505) capable of monitoring at least one network parameter associated with said data network (120);

a second controller (515) capable of monitoring in said decoder buffer (131) a delay budget region comprising a sequence of access units that are about to be accessed sequentially by said data decoder (134), said delay budget region comprising a retransmission  
10 region and a late region separated by a temporal boundary, wherein said second controller (515) detects missing data packets in said retransmission region and said late region and, in response to detection of a missing data packet in said retransmission region, transmits a retransmission request for said missing data packet to said streaming transmitter (110), and wherein said second controller (515) is capable of adjusting said temporal boundary to  
15 thereby advance or retard said transmission of said retransmission request.

2. The delay budget controller (138) set forth in Claim 1 wherein said second controller (515) adjusts said temporal boundary in response to a measured value of said at least one network parameter.

20 3. The delay budget controller (138) set forth in Claim 2 wherein said at least one network parameter comprises a round trip delay period associated with said retransmission request.

25 4. The delay budget controller (138) set forth in Claim 3 wherein said at least one network parameter comprises a delay jitter associated with a variation in transmission delay from said streaming transmitter (110) to said decoder buffer (131).



5. The delay budget controller (138) set forth in Claim 2 wherein said at least one network parameter comprises an available bandwidth value associated with a communication channel (120) between said streaming data transmitter and said decoder buffer (131).

5 6. The delay budget controller (138) set forth in Claim 2 wherein said first and second controllers (505, 515) are capable of determining a probability that a packet that is identified as lost by said first and second controllers (515) is actually lost.

10 7. The delay budget controller (138) set forth in Claim 6 wherein said second controller (515) adjusts said temporal boundary in response to a value of said probability.

15 8. The delay budget controller (138) set forth in Claim 2 wherein said second controller (515) is capable of adjusting a second temporal boundary associated with said delay budget region to thereby increase or decrease a duration of said delay budget region.

9. A video system comprising:

a decoder buffer (131) capable of receiving streaming video packets over a data network (120) from a streaming video transmitter and storing said video packets in a plurality of access units for subsequent retrieval;

20 a delay budget controller (138) comprising:

a first controller (505) capable of monitoring at least one network parameter associated with said data network (120);

25 a second controller (515) capable of monitoring in said decoder buffer (131) a delay budget region comprising a sequence of access units that are about to be accessed sequentially, said delay budget region comprising a retransmission region and a late region separated by a temporal boundary, wherein said second controller (515) detects missing video packets in said retransmission region and said late region and, in response to detection of a missing video packet in said retransmission region, transmits a retransmission request for said missing video packet to said streaming transmitter (110), and wherein said second  
30 controller (515) is capable of adjusting said temporal boundary to thereby advance or retard said transmission of said retransmission request;

a video decoder capable of retrieving said video packets from said decoder buffer (131) and decoding said video packets; and

a display device coupled to said video decoder for displaying said decoded video packets.

10. For use with a decoder buffer (131) capable of receiving streaming data  
5 packets over a data network (120) from a streaming transmitter (110) and storing the data packets in a plurality of access units for subsequent retrieval by a streaming data decoder (134), a method of controlling retransmission requests associated with lost ones of the data packets comprising the steps of:

10 monitoring at least one network parameter associated with the data network (120);

monitoring in the decoder buffer (131) a delay budget region comprising a sequence of access units that are about to be accessed sequentially by the data decoder (134), the delay budget region comprising a retransmission region and a late region separated by a temporal boundary;

15 detecting missing data packets in the retransmission region and the late region; in response to detection of a missing data packet in the retransmission region, transmitting a retransmission request for the missing data packet to the streaming transmitter (110); and

20 adjusting the temporal boundary to thereby advance or retard the transmission of the retransmission request.

11. For use with a decoder buffer (131) capable of receiving streaming data packets over a data network (120) from a streaming transmitter (110) and storing the data packets in a plurality of access units for subsequent retrieval by a streaming data decoder  
25 (134), computer-executable process steps stored on a computer-readable storage medium (140) for controlling retransmission requests associated with lost ones of the data packets comprising the steps of:

monitoring at least one network parameter associated with the data network (120);

30 monitoring in the decoder buffer (131) a delay budget region comprising a sequence of access units that are about to be accessed sequentially by the data decoder (134), the delay budget region comprising a retransmission region and a late region separated by a temporal boundary;

detecting missing data packets in the retransmission region and the late region;

in response to detection of a missing data packet in the retransmission region,  
transmitting a retransmission request for the missing data packet to the streaming transmitter  
(110); and

adjusting the temporal boundary to thereby advance or retard the transmission  
5 of the retransmission request.

1/4

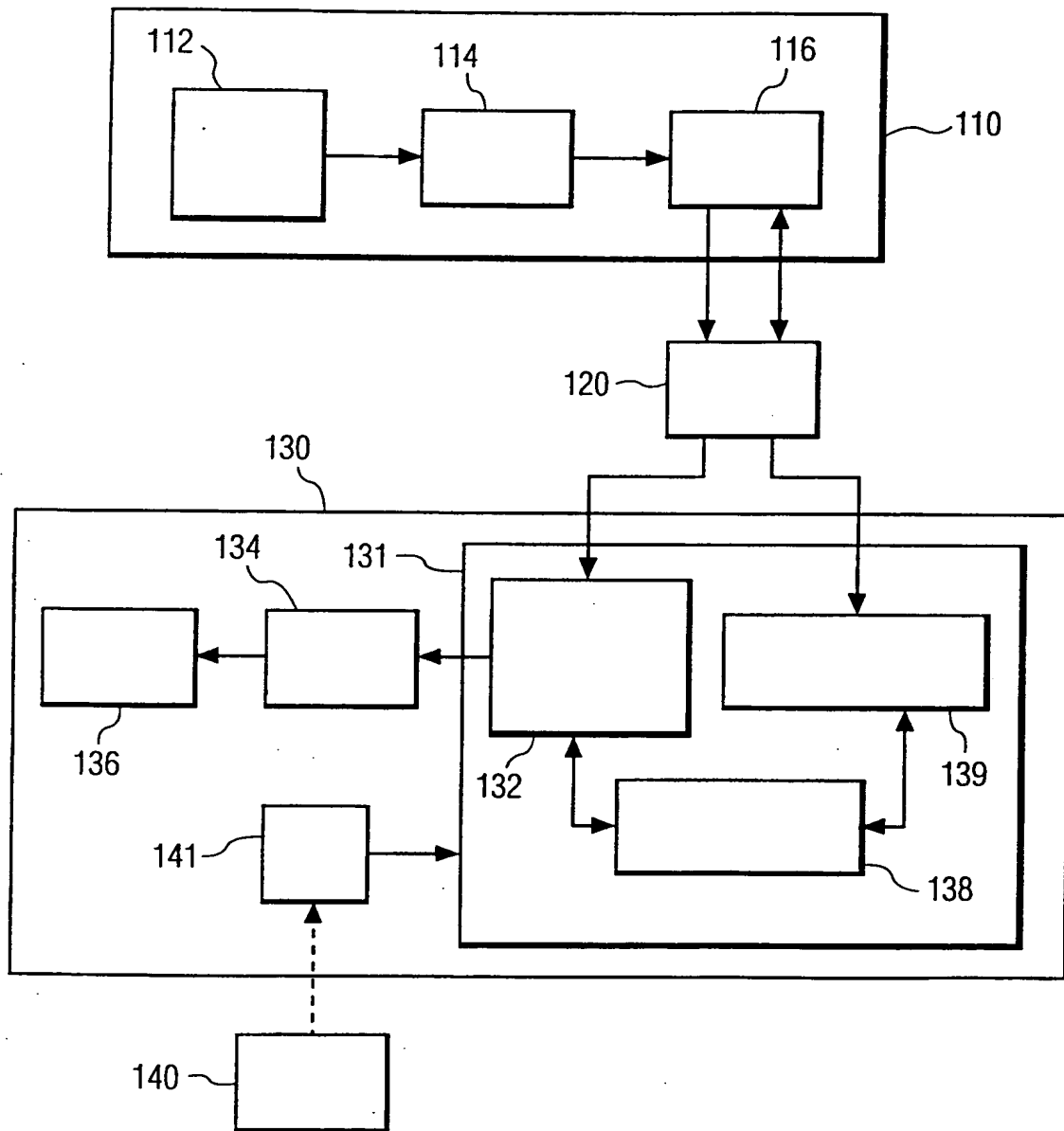


FIG. 1

2/4

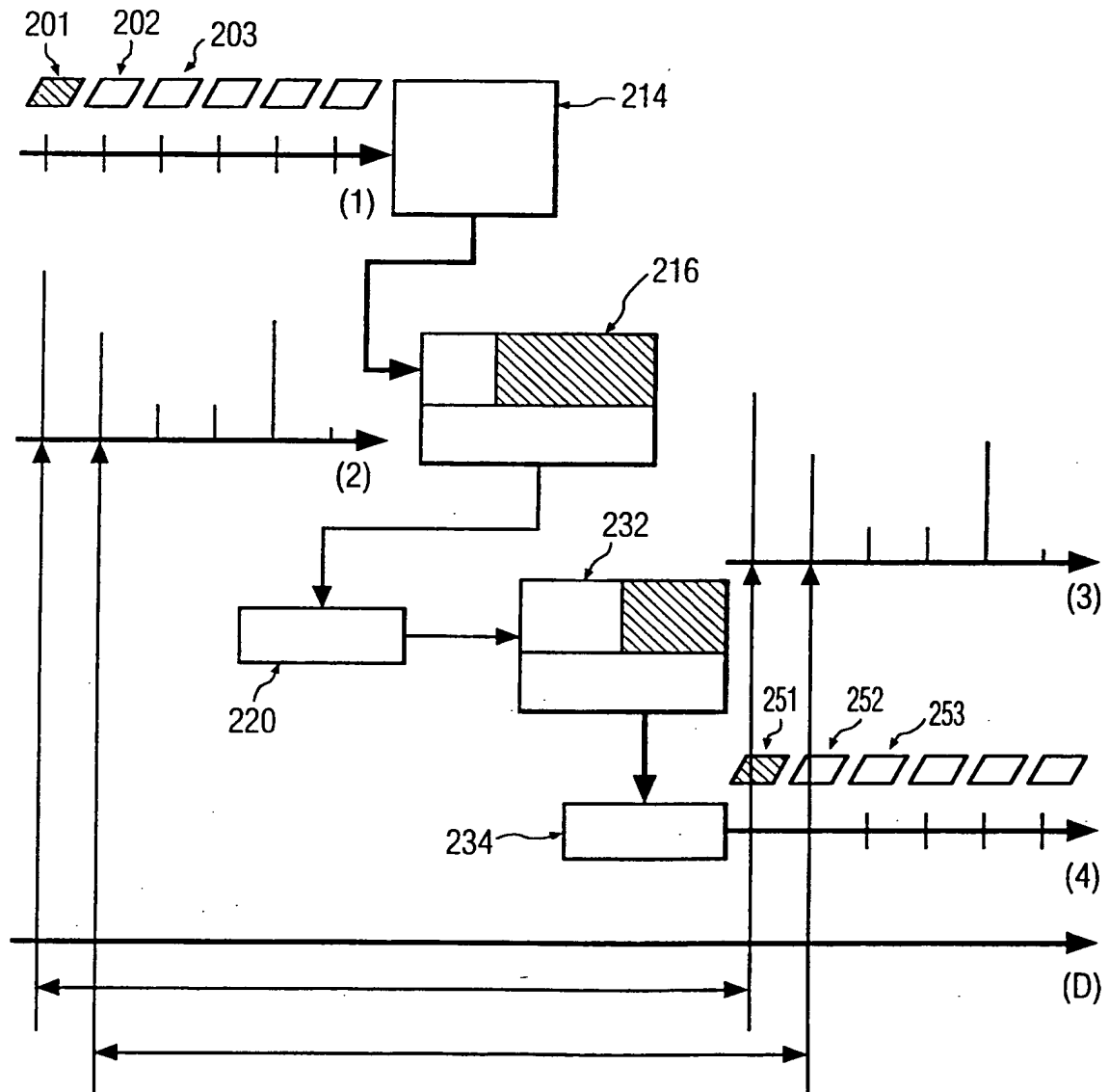


FIG. 2

3/4

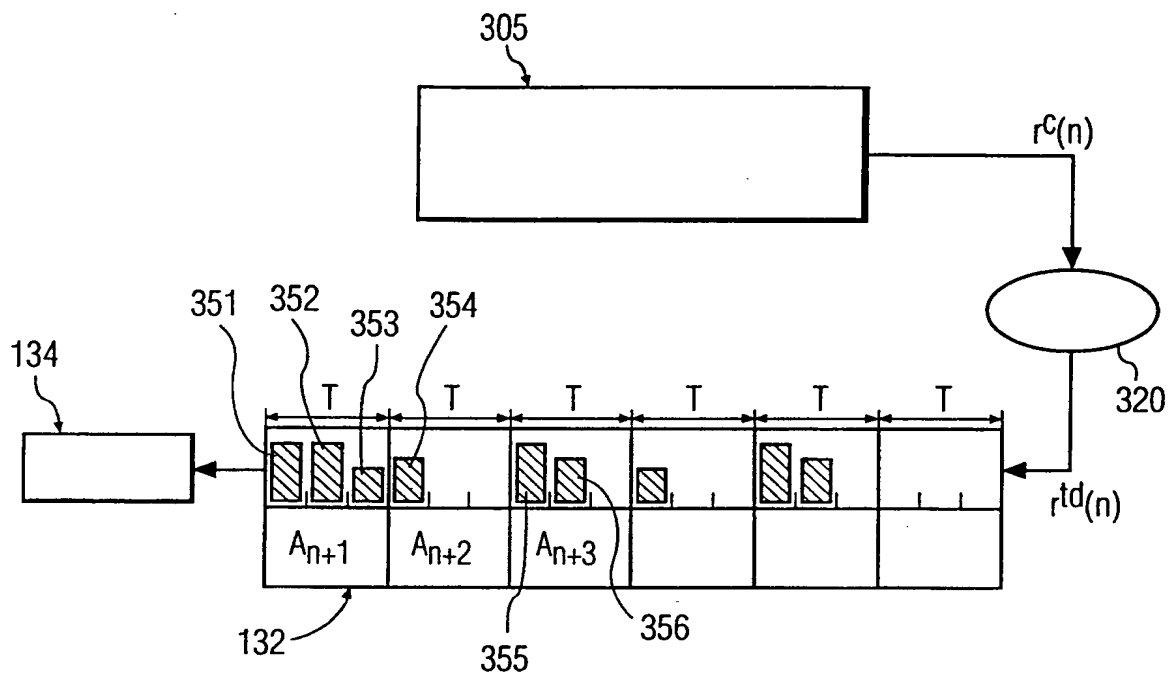


FIG. 3

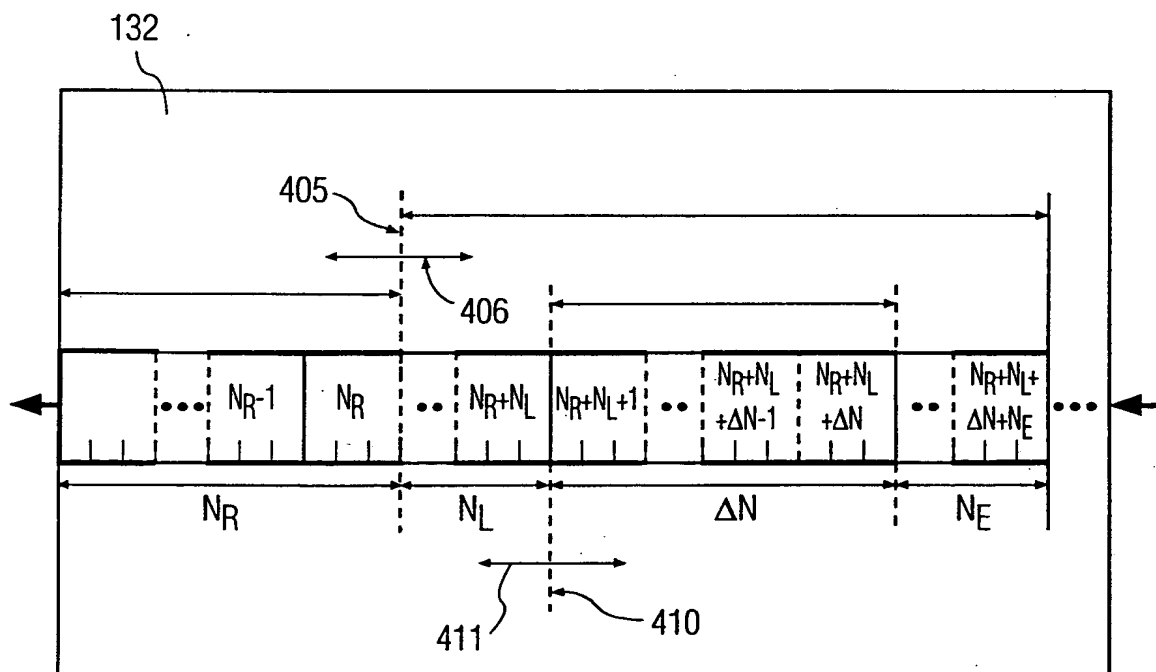


FIG. 4

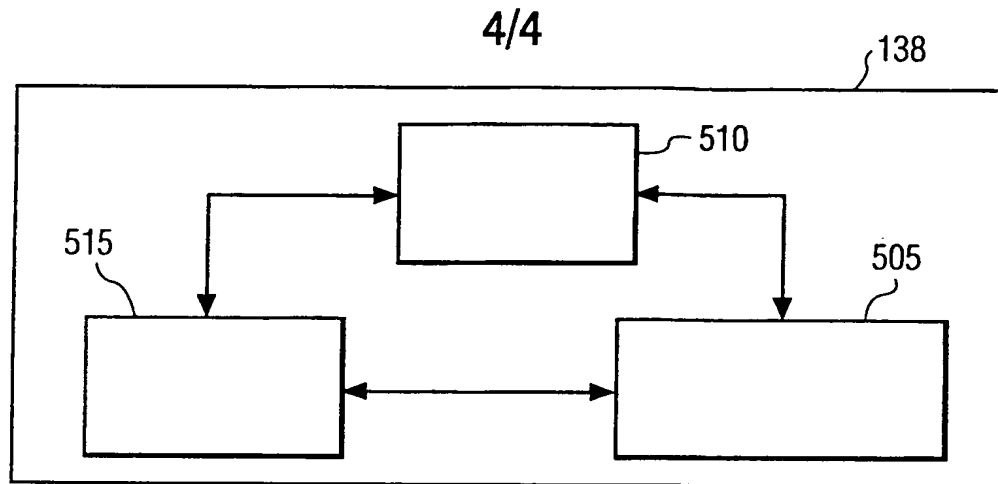


FIG. 5

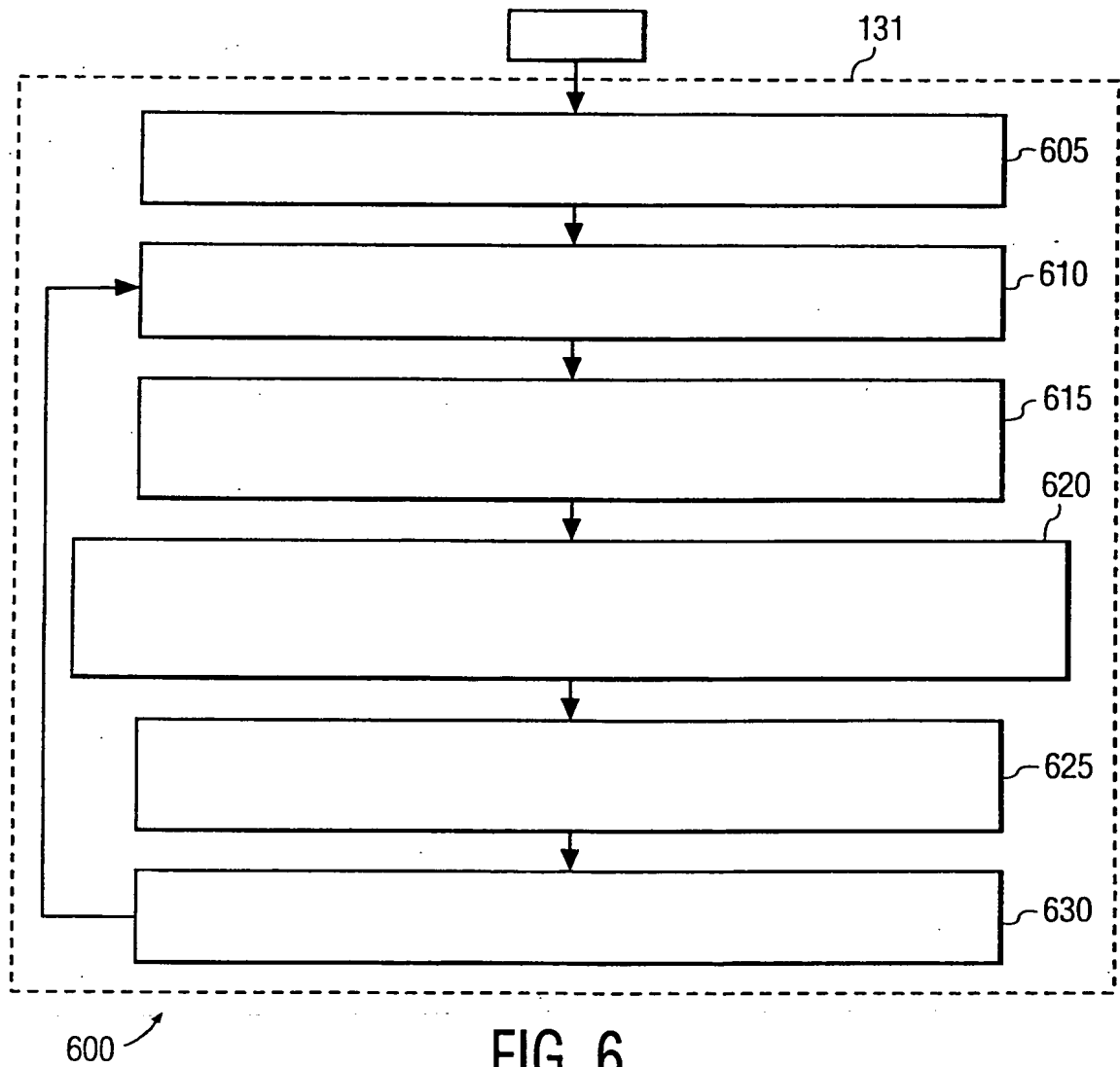


FIG. 6

# INTERNATIONAL SEARCH REPORT

Int. Application No

PCT/EP 00/10902

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 H04N7/24 H04L12/64

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 H04N H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5 768 527 A (ZHU QIN-FAN ET AL) 16 June 1998 (1998-06-16) column 2, line 5 - line 19 column 2, line 48 - line 65 column 4, line 4 - line 28 column 5, line 21 - line 24 column 7, line 4 - line 48 figure 6	1-11
Y	DAVIC: "Davic 1.5 Specification, jitter concealment tools, revision 1.0" 22 January 1999 (1999-01-22), DAVIC (DIGITAL AUDIO VISUAL COUNCIL), PAGES I-II AND 1-9 XP002162802 page 7-8, paragraph 8.2.4 figure 8.4	1-11
	--- -/--	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*&\* document member of the same patent family

Date of the actual completion of the international search

20 March 2001

Date of mailing of the international search report

04/04/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Fantini, F



# INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 00/10902

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>US 5 623 483 A (AGRAWAL PRATHIMA ET AL)  22 April 1997 (1997-04-22)  abstract  column 1, line 61 -column 2, line 20  column 2, line 65 -column 3, line 19  column 4, line 12 - line 63  column 6, line 7 - line 19</p>	1-11
A	<p>US 5 918 002 A (RAVI HEMANTH SRINIVAS ET AL)  29 June 1999 (1999-06-29)</p>	
A	<p>WO 97 22201 A (XIE DONG ;CAMPBELL ROY H (US); CHEN ZHIGANG (US); TAN SEE MONG (US)  19 June 1997 (1997-06-19)</p>	
A	<p>WO 95 22233 A (BESSETTE FRANCOIS ;NEWBRIDGE NETWORKS CORP (CA))  17 August 1995 (1995-08-17)</p>	

# INTERNATIONAL SEARCH REPORT

Int. l. Application No  
PCT/EP 00/10902

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5768527 A	16-06-1998	US 6085252 A	04-07-2000
US 5623483 A	22-04-1997	NONE	
US 5918002 A	29-06-1999	NONE	
WO 9722201 A	19-06-1997	EP 0867003 A JP 2000515692 T	30-09-1998 21-11-2000
WO 9522233 A	17-08-1995	AU 1572995 A	29-08-1995